

Julia Schütz, Lena Rosenkranz, Anne Martin

Evaluation –

ausgewählte und kommentierte Texte

Fakultät für
**Kultur- und
Sozialwissen-
schaften**

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Der Inhalt dieses Studienbriefs wird gedruckt auf Recyclingpapier (80 g/m², weiß), hergestellt aus 100 % Altpapier.

Inhaltsverzeichnis

Inhaltsverzeichnis.....	3
Vorwort.....	4
Zusammenfassung.....	5
Lernziele.....	7
1 Kuper, Harm. Evaluation.....	8
2 Stockmann, Reinhard & Meyer, Wolfgang. Wissenschaftsbasierte Evaluation	20
3 Stockmann, Reinhard & Meyer, Wolfgang. Evaluationsprozess.....	73
4 Stockmann, Reinhard & Meyer, Wolfgang. Informationssammlung und -bewertung	107
5 Flick, Uwe. Qualitative Methoden in der Evaluationsforschung.....	168
6 Dahler-Larsen, Peter. Die sich verändernde Rolle der Evaluation in einer sich verändernden Gesellschaft.	179
Hinweise zu den Autor*innen	193
Kommentierte Bibliographie.....	195
Glossar	196
Literaturverzeichnis	197
Internetquellen	197

Vorwort

In diesem Kurs finden Sie ausgewählte wissenschaftliche Beiträge über die Evaluationsforschung. Durch das Studium der Texte und die Bearbeitung der Reflexionsfragen bekommen Sie einen Überblick über zentrale Begriffe, Theorien und Diskurse. Außerdem stellen die Texte eine umfassende, praktische Hilfestellung bei der Durchführung Ihrer eigenen Evaluationsstudie dar. Evaluationen sind stets eingebettet in politische, institutionelle, pädagogische, soziologische und methodische Kontexte. Diesen kontextuellen Bedingungen kommt in den Beiträgen eine besondere Aufmerksamkeit zu und erhöht Ihr Verständnis für spezifische Rahmenbedingungen von Evaluationen.

Der erste Text aus dem Sammelband „Empirische Bildungsforschung“, herausgegeben von Heinz Reinders, Hartmut Ditton, Cornelia Gräsel und Burkhard Gniewosz, führt in die Grundbegriffe und methodischen Herausforderungen der Evaluationsforschung ein. Es folgen drei Beiträge aus dem Buch mit dem Titel „Evaluation. Eine Einführung“ von Reinhard Stockmann und Wolfgang Meyer. In diesen wird das Spannungsverhältnis von wissenschaftlichen und politischen Belangen bei einer Evaluation ausführlich bestimmt (Kapitel 2) und überführt in eine prozessuale und funktionale Betrachtung von Evaluationen (insbesondere Kapitel 3). Mit dem vierten Kapitel bekommen Sie eine Anleitung zur Erhebung, Aufbereitung und Auswertung von evaluationsbasierten Daten. Ausführlicher wird im anschließenden Text aus der Zeitschrift für Qualitative Forschung von Uwe Flick die qualitative Evaluationsforschung in den Blick genommen. Schließlich folgt ein Beitrag aus dem Buch „Die Zukunft der Evaluation“ von Peter Dahler-Larsen, in dem ein konstruktivistisches Konzept von Evaluationen vorgestellt und mit zentralen gesellschaftlichen Veränderungsprozessen in Zusammenhang gebracht wird.

Die Texte wurden von Prof. Dr. Julia Schütz, Dr. Lena Rosenkranz und Anne Martin, M.A., am Lehrgebiet Empirische Bildungsforschung der FernUniversität in Hagen zusammengestellt, kommentiert und mit Reflexionsfragen versehen. Die Beiträge der verschiedenen Autor*innen wurden unverändert übernommen. Alle Kennzeichnungen in dieser Form [OV, S.] verweisen auf die Seitenzahlen der Originalversion. Bitte verwenden Sie diese Seitenangaben beim Zitieren in Ihren wissenschaftlichen Arbeiten. Zu den einzelnen Beiträgen wurden Fragen zum Weiterdenken formuliert, die auch in der Online-Lehre Berücksichtigung finden und gemeinsam in den Foren diskutiert werden können. Den Beiträgen sind jeweils eine kurze Zusammenfassung sowie eine Sammlung von Schlüsselwörter vorangestellt. Ebenfalls befinden sich ausformulierte Lernziele zu Anfang des Studienbriefs. Eine kommentierte Bibliographie und ein speziell auf vorliegenden Beiträge ausgerichtetes Glossar schließen den Studienbrief ab.

Wir wünschen Ihnen viel Erfolg bei der Bearbeitung und eine anregende Lektüre.

Zusammenfassung

In diesem Kurs finden Sie ausgewählte Texte, die Ihnen die Grundlagen der Evaluationsforschung vermitteln. Sie lernen evaluationsspezifische Verfahren, Begriffe und Modelle kennen, setzen sich mit Potentialen, Herausforderungen sowie Limitationen auseinander und sensibilisieren sich für die sozialen, institutionellen, politischen und wissenschaftlichen Rahmenbedingungen von Evaluationen. Detaillierte Prozessbeschreibungen können Sie als Leitfaden für Ihre eigenen Forschungsprojekte nutzen.

Textübergreifend werden zentrale Diskussionslinien in der Evaluationsforschung offenbar. Alle Beiträge umkreisen die Frage nach der Qualität von Evaluationen. Dahinter verbirgt sich zum einen eine grundlegende Frage: Wie können wir angesichts der komplexen und multiplen Untersuchungsgegenstände eine methodisch kontrollierte und gegenstandsnahe Datenbasis schaffen? Dieser Anspruch ist bei Evaluationen von besonderer Relevanz, weil sie immer eine Bewertung oder Beurteilung implizieren, die dann auch Auswirkungen auf die Praxis haben (können). Damit stellt sich zum anderen die Frage, wie wir möglichst objektive Bewertungen kreieren, die erstens an alle Beteiligten vermittelt werden können, zweitens an der Praxis anknüpfen und drittens Nachhaltigkeit anvisieren. Evaluationen bewegen sich im besonderen Maße im Spannungsfeld zwischen Politik, Wissenschaft und Praxis sowie im Spannungsfeld zwischen Subjektivität und Objektivität. Die folgenden Beiträge liefern Ihnen genauere Beschreibungen der Problematiken und thematisieren immer auch forschungspraktische Wege im Umgang mit diesen Herausforderungen.

Harm Kuper, Professor für Weiterbildung und Bildungsmanagement an der Freien Universität Berlin, veranschaulicht das Potential von Evaluationen anhand des Bildungssystems. Das Forschungsfeld ist umfassend: Institutionen, Organisationen, Curricula, Unterrichtsstunden, Projekte, Lehrmaterialien u.a. werden von Evaluationsforscher*innen in den Blick genommen und ziehen zahllose Forschungsfragen mit sich. Harm Kuper rückt insbesondere die Untersuchung von Leistungsvergleichen in den Fokus: Wie lassen sich verschiedene Gruppen, Programme oder Systeme hinsichtlich ihrer Leistungen vergleichen und wie nehmen wir Leistungsentwicklungen in den Blick? Er schließt die Frage an, welche Aussagekraft die gewonnenen Ergebnisse für die Qualität des Bildungssystems haben und wo sie uns hinführen. Der Beitrag führt in die Evaluationsforschung ein und schließt inhaltlich an den Studienbrief zur „Empirischen Bildungsforschung“ an.

Die folgenden drei Beiträge sind geschrieben von Reinhard Stockmann und Wolfgang Meyer (Centrum für Evaluation: CEval). Die Autoren verdeutlichen die Sonderstellung von Evaluationen als anwendungsbezogene Forschung, indem sie das Spannungsfeld zwischen Evaluation und Politik erörtern. Evaluationen sind oft Auftragsarbeiten und bergen somit potenziell Interessenskonflikte, eine begrenzte Vertrautheit mit dem Forschungsgegenstand, richtungsweisende institutionelle Rahmenbedingungen, begrenzte Ressourcen u.a. Die Autoren veranschaulichen die Spannungen anhand der forschungsleitenden Fragen: Was wird wozu, anhand welcher Kriterien, von wem und wie evaluiert? (vgl. Stockmann & Meyer 2014, S. 75). Sie stellen als Antwort auf die Fragen jeweils weiterführende Begriffe, Modelle und Möglichkeiten im Rahmen einer Evaluation vor und liefern somit forschungspraktische Hinweise für den eigenen Forschungsprozess. Im

Kapitel „Evaluationsprozess“ gehen die Autoren noch einmal detaillierter auf den Ablauf einer Evaluation ein. Schritt für Schritt werden die Parameter bei der Planung, Durchführung und Verwertung beschrieben und kontextualisiert. Der Berücksichtigung von Qualitätsstandards in der Evaluationsforschung wird dabei besondere Aufmerksamkeit gewidmet. Als verbindendes Modell stellen Stockmann und Meyer den Evaluationsansatz des CEvals vor, der Forschungsphasen, Aufgaben und Akteure systematisch aufführt. Im Kapitel „Informationssammlung und -bewertung“ werden nun noch einmal methodische Verfahrensweisen diskutiert. Es gibt einen Überblick zu qualitativen und quantitativen Datenerhebungsmethoden und auszugsweise werden einige Auswertungsverfahren vorgestellt. In dem Kapitel werden wir für Fehlerquellen und Rahmenbedingungen sensibilisiert. Integrierender Ansatzpunkt der Autoren ist der partizipative Evaluationsprozess: ein Verfahren, das die Stakeholder systematisch einbindet und wissenschaftliche Standards impliziert.

Der Beitrag von Uwe Flick vertieft noch einmal den Fokus auf qualitative Forschungsmethoden bei Evaluationen und nimmt eine Standortbestimmung vor. Gerade dieser Beitrag schließt an den partizipativen Ansatz von Stockmann und Meyer an, denn der Berücksichtigung der Perspektiven aller Beteiligten kommt in der qualitativen Forschung ein besonderer Stellenwert zu. Die Qualitätsmerkmale von Evaluationen und die Gütekriterien der qualitativen Forschung korrelieren und stellen somit eine sinnvolle Ergänzung oder Alternative zu quantitativen Verfahren dar. Expert*inneninterviews, Gruppendiskussionen und Vignetten werden von Uwe Flick hervorgehoben und der Wert dieser methodischen Verfahren für die Evaluationsforschung markiert.

Peter Dahler-Larsen befasst sich mit Evaluationen aus einer konstruktivistischen Perspektive und erweitert dadurch noch einmal den Blick auf zentrale Konflikt- und Problemthemen im Spannungsfeld zwischen Wissenschaft, Politik und Praxis. Was machen wir, wenn die Standards, Empfehlungen und Erfahrungen von Evaluator*innen – wie wir sie in den anderen Texten kennengelernt haben – nicht berücksichtigt werden? Was machen wir mit Evaluationen, die ungenutzt bleiben, es sich also um eine Qualitätsmessung handelt, die keine Verbesserung der Qualität zur Folge hat (vgl. Dahler-Larsen 2018, S. 32)? Mit diesen und weiteren kritischen Fragen befasst sich Peter Dahler-Larsen in seinem Beitrag.

Zusammenfassend beinhaltet der Kurs zum einen praxisorientierte, systematische Handlungsanleitungen für die Planung, Durchführung und Umsetzung von Evaluationsprojekten. Zum anderen rekurren die Texte aus unterschiedlichen Perspektiven auf die zentralen Diskurslinien der Evaluationsforschung und sensibilisieren damit für die Herausforderungen, die sich während des Evaluationsprozesses stellen. Es sollen diese Inhalte sein, die im Modul „Anwendungsbezogene Bildungsforschung“ Diskurse entfachen und zu gewinnbringenden Evaluationen ermutigen.

Lernziele

Mit der Bearbeitung dieses Kurses sind folgende Lernziele verbunden:¹

- Sie können Evaluation definieren und ihre Besonderheiten benennen.
- Sie haben sich mit den Funktionen und Zielen von Evaluationen vertraut gemacht.
- Sie kennen Ansätze der Evaluationsforschung und ihre historische Entwicklung.
- Sie können verschiedene methodische und theoretische Zugänge in der Evaluationsforschung beschreiben.
- Sie können die kontextuellen Bedingungen einer Evaluation analysieren und methodisch integrieren.
- Sie haben sich mit verschiedenen Evaluationsstudien auseinandergesetzt und können diese hinsichtlich ihrer Limitationen beurteilen.
- Sie kennen die Prozessschritte einer Evaluation und sind sich verschiedener Fehlerquellen bewusst.
- Sie sind mit den Qualitätsstandards der Deutschen Gesellschaft für Evaluation (DeGEval) sowie mit weiteren Qualitätskriterien vertraut.
- Sie sind in der Lage, ein eigenes Evaluationsprojekt zu planen und durchzuführen.
- Sie reflektieren über das Spannungsverhältnis zwischen Wissenschaft und Praxis sowie zwischen Objektivität und Subjektivität.
- Sie verstehen die konstruktivistische Perspektive auf Evaluationen und kennen zentrale Diskussionslinien in der Evaluationsforschung.

¹ Die Lernziele orientieren sich an der Lernzieltaxonomie nach Bloom (1956).

1 Kuper, Harm (2015). Evaluation

Kuper, Harm (2015). Evaluation In: Reinders, Heinz, Ditton, Hartmut, Gräsel, Cornelia & Gnievosz, Burkhard (Hrsg.) Empirische Bildungsforschung. Strukturen und Methoden, 2., überarbeitete Auflage. Wiesbaden: Springer VS, S. 141- 153.

Zusammenfassung durch das Lehrgebiet

Harm Kuper ist Lehrstuhlinhaber der Professur für Weiterbildung und Bildungsmanagement an der Freien Universität Berlin. Er nähert sich im vorliegenden Text dem Begriff „Evaluation“ zunächst aus einer alltagsnahen Perspektive: Wir alle sind vertraut mit dem grundlegenden Prinzip der Evaluation, bei dem es zunächst um das Bewerten sowie Abwägen eines Sachverhalts geht, um davon ausgehend zukunftsgerichtete Schlüsse zu ziehen. Der Autor grenzt darauf aufbauend, die wissenschaftlichen Prinzipien einer Evaluation von dieser alltagstauglichen Perspektive ab, indem er die Standards und Funktionen einer Evaluation unter Beachtung zentraler Gütekriterien beschreibt. Anschließend wird die Evaluationsforschung in der Mehrebenenstruktur des Bildungssystems betrachtet. Eine Herausforderung besteht in diesem Forschungsfeld darin, bei der Aggregation von Daten, den Informationsgewinn mit dem -verlust abzuwägen und methodisch zu kontrollieren. Dabei spielt es immer auch eine Rolle, welche Teile des Bildungssystems und welche Einflussfaktoren in den Fokus gerückt werden. Dazu finden sich im Text Ausführungen zum Bildungsmonitoring als vergleichende Beschreibung der Leistungsfähigkeit in Bildungssystemen ebenso wie zu Lernstandserhebungen und Programm- oder Projektevaluationen als Anwendungsbereiche von Evaluationen. Weiterhin stellt der Artikel besondere Methodenprobleme der Evaluation in Bezug auf die Methode des Experiments oder des Vergleichs vor. Diese beziehen sich im ersten Fall auf die Kontrolle der Untersuchungssettings und die Rückführung von Effekten auf das Treatment der Untersuchungspersonen sowie im zweiten Fall auf den Einbezug von Leistungsvoraussetzungen bei der Beurteilung von Leistungen. Der letzte Abschnitt des Artikels widmet sich der Bewertung von Evaluationsergebnissen, die von den drei Größen einer Evaluation abhängt: ihrer wissenschaftlichen Güte, der Ergebnisvermittlung und ihres Transfers in die Praxis.

Schlüsselwörter

Definition der Evaluation • Gütekriterien in der Evaluation • Evaluationsforschung • Methodenprobleme der Evaluation • Verwendung von Evaluationsergebnissen

Zusammenfassung [durch den Autor]: Grundbegriffe der Evaluation werden eingeführt und Linien der Diskussion über Evaluationskonzepte skizziert. Im Zentrum steht Evaluation als angewandte Forschung. Davon ausgehend werden Konzepte der Evaluationsforschung im Bildungswesen, besondere methodische Anforderungen der Evaluationsforschung sowie Perspektiven der Verwendung von Evaluationsergebnissen aufgezeigt.

1 Allgemeines und Definition

Evaluation ist – in einem naiven Sinne – in vielen Situationen des praktischen Handelns gegenwärtig. Praktisches Handeln bedeutet, Entscheidungen zu treffen. Etwa: Fahre ich besser mit dem Bus oder mit dem Fahrrad zur Universität? Idealtypisch lassen sich Entscheidungen in zwei Komponenten zerlegen. Sie haben erstens einen normativen Anteil, d. h. Entscheidungen sollen einen als günstig beurteilten Zustand zur Folge haben. In dem gewählten Beispiel ist die Dauer der Fahrt zur Universität ein denkbare normatives Kriterium. Zweitens haben Entscheidungen einen „technologischen“ Anteil, d. h. es soll aus einer Alternative verfügbarer Möglichkeiten die im Sinne des normativen Kriteriums beste ausgewählt werden. Ich entscheide mich also für das Fahrrad, weil die Fahrt erwartungsgemäß nur 10 Minuten dauert, ich aber nach Fahrplan 8 Minuten warten müsste, um eine voraussichtlich 4-minütige Busfahrt anzutreten. Nun sind aber alle Entscheidungen in normativer und technologischer Hinsicht riskant. So kann ich zu der Einsicht gelangen, dass Bequemlichkeit ein wichtigeres Urteilkriterium ist als Schnelligkeit. Ich kann erleben, dass der Bus mich trotz verspäteter Abfahrt vor dem Ziel überholt, weil ich den Gegenwind unterschätzt habe usw. Weil also in der Entscheidungssituation keine vollständige Kontrolle über die Folgen der Entscheidung gegeben ist, bleibt [OV, S. 141] nur die Möglichkeit, nach erfolgtem Ablauf rückblickend den gesamten Vorgang und sein Ergebnis zu evaluieren – d. h. angesichts der ursprünglichen und eventuell zwischenzeitlich hinzugekommenen Kriterien zu bewerten, ob es wirklich günstig ist, nach 17 Minuten abgehetzt an der Universität anzukommen, obgleich man es mit dem Bus bequem in 15 Minuten geschafft hätte – und daraus für weitere Entscheidungen praktische Konsequenzen zu ziehen: Ab morgen fahre ich Bus!

Zwei Komponenten von Entscheidungen

- **Evaluation heißt somit: Lernen aus Erfahrung, um für zukünftige Entscheidungen praktisch verwertbare Informationen und normativ tragfähige Urteilkriterien zu erhalten.**

Auch einem wissenschaftlichen Begriffsverständnis zufolge bezieht sich Evaluation auf eine normative und eine technologische Dimension von Entscheidungen (Kuper, 2005). Gegenstände der Evaluation sind dann aber von einer weitaus größeren Komplexität und Tragweite, als das normalerweise bei alltäglichen Entscheidungen der Fall ist. So können im Bildungssystem Institutionen und Organisationen, Curricula, Unterrichtsstunden, Projekte, Lehrmaterialien und vieles mehr einer Evaluation unterzogen werden. Auch diese Gegenstände gehören einer Praxis an und sind von riskanten Entscheidungen abhängig, die normative Urteilsfähigkeit sowie sachliche Informationen über die zu erwartenden Folgen voraussetzen. Der Bezug auf Handeln bzw. Entscheidung ist ein bedeutendes Merkmal der Evaluation. Auf ihn wird in Definitionen übereinstimmend hingewiesen. So bestimmt Suchman (1971, S. 44) den Gegenstandsbereich der Evaluation als

Wissenschaftliches Begriffsverständnis

„(1) a planned program or deliberate intervention, not just any natural or ‚accidental‘ event; (2) an objective or goal which is considered desirable or has some positive value, not simply whatever change occurs; and (3) a method for determining the degree to which the planned program achieves the desired objective.“

Aspekte wissenschaftlicher Evaluation

Mertens (1998, S. 219) definiert fast 30 Jahre später: „Evaluation is the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision making.“ Hervorzuheben ist der übereinstimmende Gebrauch der Vokabel „program“ in beiden Definitionen. Er findet

sich auch in dem Begriff der Programmevaluation. Darüber hinaus gehen beide Definitionen von einer Methodik bzw. Systematik in der Evaluation aus, die durch die Verwendung wissenschaftlicher

Evaluation als die Kunst des Möglichen

Verfahren der Datenerhebung und -auswertung verbürgt ist. Der Aspekt der Wissenschaftlichkeit wird in der Bezeichnung Evaluationsforschung betont. Der wissenschaftliche Charakter der Evaluation ist ein stetiger Quell der Diskussion. Einerseits besteht die Ansicht, Evaluation(-sforschung) bediene sich der

gleichen Methoden, wie jede andere wissenschaftliche [OV, S. 142] Forschung auch und sei lediglich durch den engen Anwendungsbezug unterscheidbar. Diese Position ging über lange Zeit mit der Überzeugung einher, allein die wissenschaftliche Güte der Evaluation sei ausschlaggebend, um aus ihr unmittelbar praktische Entscheidungen – inklusive normativer Beurteilung – ableiten zu können. Dieser Ansicht hing z. B. Scriven (1971, S. 53) an, der von einem Evaluator erwartete: „It’s his task to try very hard to condense all the mass of data into one word: good, or, bad“. Auf der anderen Seite geht die Erwartung der praktischen Relevanz von Evaluationsstudien einher mit einer kritischen Einstellung gegenüber einem linearen Transfer von Evaluationsresultaten in praktische Entscheidungen. Dieses beschäftigte den Methodologen Cronbach (1982, S. 321), als er formulierte: „Evaluation ist die Kunst des Möglichen“. Er vertritt ein Konzept der Evaluation, in dem forschungsmethodische Ansprüche pragmatisch mit den situativen Möglichkeiten der Realisierung von Forschungsdesigns sowie mit den Informationsbedürfnissen in den Anwendungskontexten zu integrieren sind.

Standards und Funktionen der Evaluation

Gütekriterien für Evaluationen

Die systematische Diskussion um Evaluation hat in vielfältiger Weise an der Position Cronbachs angeschlossen und diese ausdifferenziert. So sind für die Evaluation Gütekriterien entwickelt worden, die nicht nur die wissenschaftliche

Qualität der Datenerhebung und -auswertung betreffen, sondern auch insbesondere an der praktischen Verwendbarkeit und Verantwortung der Evaluation ausgerichtet sind.

Merksatz

Die Gütekriterien sind zusammengefasst in den Gruppen der Nützlichkeits-, Durchführbarkeits-, Korrektheits- und Genauigkeitsstandards (Kuper, 2005, S. 114 ff.).

Die Titel der Standardgruppen lassen erahnen, dass die Güte der Evaluation sich nicht nur am Leitmotiv der Erkenntnis bemisst, sondern zu erheblichen Anteilen an Erwartungen der Urteils- und Handlungsfähigkeit eines praktisch verantwortlichen Akteurs, der methodisch kontrollierte Wissensproduktion mit Sensibilität für den Informationsbedarf der Praxis verbindet. Dem korrespondieren auch die Funktionen, die der Evaluation zugeschrieben werden. Neben *Erkenntnis* identifiziert Stockmann (2000) *Legitimation*, *Kontrolle* und *Dialog* als Funktionen der Evaluation. Diese

bezeichnen gleichzeitig Leistungen, die von Evaluation im Kontext von Management, Politik und Programmentwicklung erwartet werden.

Die Besonderheiten der Evaluation können in Abgrenzung zur Grundlagenforschung nicht primär über die Forschungsmethoden bestimmt [OV, S. 143] werden – sofern es überhaupt evaluationsspezifische Methoden gibt, sind diese allenfalls als Varianten aus dem Spektrum der qualitativen und quantitativen Verfahren zu beschreiben. Vielmehr ist die Kopplung an konkrete Verwertungsinteressen entscheidend. Daher ist es nicht verwunderlich, dass die soziale Einbettung in den Verwertungszusammenhang ein Hauptgegenstand der Reflexion über Evaluation ist. Evaluation ist in der Regel Auftragsforschung. Dabei können die Auftraggeber das zu evaluierende Programm selbst verantwortlich betreiben oder als Finanziere Leistungen von einem Programm erwarten. Evaluationsergebnisse richten sich an unterschiedliche Gruppen von Adressaten. Neben den Auftraggebern – die im Regelfall auch Rezipienten der Evaluationsergebnisse sind – können unter anderem die Öffentlichkeit, die Veranstalter kooperierender und bisweilen auch konkurrierender Programme, Mitarbeitende, Politiker, Teilnehmer und Interessenten eines Programms Nutzer der Evaluationsergebnisse sein. Bei der Vielzahl der unterschiedlichen „stakeholder“ – so Formative und summative Evaluation werden im Jargon der Evaluation diejenigen genannt, die von der Durchführung der Evaluation betroffen sind – liegt es nahe, dass Evaluation mit Interessensdivergenzen, Wertepluralismus und unterschiedlichen Handlungsrationalitäten konfrontiert wird. Diese lassen sich zwar bedingt im Rahmen der Evaluation durch gut gewählte Forschungsmethoden berücksichtigen, aber sicherlich nicht selbst unter forschungsmethodische Kontrolle bringen.

Evaluation als Auftragsforschung

Eine ebenso verbreitete wie hilfreiche erste Unterscheidung der Funktionen von Evaluation im Hinblick auf den Verwertungszusammenhang ist die zwischen *formativer* und *summativer* Evaluation. Sie wird gerne mit dem folgenden, Robert Stake zugeschriebenen, Bonmot erläutert: „When the cook tastes the soup it is formative evaluation and when the guest tastes the soup it is summative.“ Hier wird deutlich, dass Evaluation einerseits als formative Variante der Verbesserung und Entwicklung eines Programms dienen kann; sie richtet sich dann vornehmlich an strategisch und operativ für das Programm verantwortliche Personen. Andererseits kann sie als summative Evaluation einen abschließenden Bericht vorlegen, der eine zusammenfassende Beurteilung eines Programms beinhaltet; in dieser Funktion ist Evaluation insbesondere für diejenigen von Interesse, die ein Programm finanzieren oder aus ihm Leistungen beziehen.

Formative und summative Evaluation

2 Evaluationsforschung in der Mehrebenenstruktur des Bildungssystems

Das Spektrum der anwendungsorientierten bildungswissenschaftlichen Forschung hat sich infolge der Auseinandersetzung mit den PISA-Resultaten (Deutsches PISA-Konsortium, 2000) und im Rahmen gesteigerter Erwartungen an die Qualitätssicherung im Bildungssystem erheblich ausdifferenziert. Ein ursprünglich eng auf Programme und Curricula bezogenes [OV, S. 144] Verständnis von Evaluation hat sich erweitert und bezieht vielfältige Forschungsaktivitäten mit evaluativer Grundhaltung ein (Baumert, 2001).